

# **42nd Symposium on the Interface of Computing Science and Statistics 2011 (Interface 2011)**

## **Statistical, Machine Learning, and Data Visualization Algorithms**

**Abstracts**

**Cary, North Carolina, USA  
1-3 June 2011**

**ISBN: 978-1-62276-714-4**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2011) by the Interface Foundation of North America  
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Interface Foundation of North America  
at the address below.

Interface Foundation of North America  
PO Box 7460  
Fairfax Station, VA 22039

Phone: (703) 993-1212  
Fax: (703) 993-1700

[interface@galaxy.gmu.edu](mailto:interface@galaxy.gmu.edu)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Program Book

## Schedule for Interface 2011

Wednesday, June 1, 2011

7:30 am – 5:00 pm                      Registration                      *Bldg. C Lobby*

8:00 am – 9:45 am

*Welcoming:* John Sall, Edward Wegman

*Keynote Address:*

21<sup>st</sup> Century Analysis, Biomedical, Collaboration and Determinism Dogma: Emerging Challenges and Guidance for Interface of Computing and Statistics  
Arnold Goodman, Collaborative Data Solutions                      *Auditorium*

9:45 am – 10:15 am                      *Morning Break*                      Prefunction Hall

10:15 am – 12:00 noon                      *Technical Sessions*

*Statistical Analysis of Attributed Graphs and Networks.*                      *Room 1012*  
Organizer: David Marchette, Naval Surface Warfare Center, Dahlgren Division

1. *Tweets to Hypergraphs*, Elizabeth Hohman, Naval Surface Warfare Center, Dahlgren Division
2. *Vertex Nomination: Finding Similar Vertices in Communications Graphs Attributed with Human Language*, Glen Coppersmith, Johns Hopkins University
3. *Combining Geographic Information in Random Dot Product Graphs*, David Marchette, Naval Surface Warfare Center, Dahlgren Division

*Predictive Modeling.* Organizer: Georgiy Bobashev,                      *Room 1014*  
Research Triangle Institute

1. *Using Random Forest Models to Predict Organizational Violence*, Burton Levine, Research Triangle Institute
2. *Patterns of Daily Alcohol Consumption*, Dan Liao, Research Triangle Institute
3. *Predictive Modeling Processes and Heavy Alcohol Use*, Barry Eggleston, Research Triangle Institute
4. *Using Tree-Based Prediction Models for Missing Data in the Complex Survey Setting*, Daryl Creel, Research Triangle Institute

12:00 noon – 1:30 pm                      Lunch Break                      *Bldg. C Café*

1:30 pm – 3:15 pm                      Technical Sessions

**Case Studies in Visualization.** **Room 1012**  
Organizer: David W. Scott, Rice University

1. *Response Surface Methodology*, John Sall, JMP/SAS Institute
2. *Exact and Approximate Area-proportional Circular Venn and Euler Diagrams*, Leland Wilkinson, SYSTAT Software, Inc.
3. *Two Micromap Extensions: 3-Way Hypothesis Test, Conditioned Micromaps and More Options for Dynamic Comparative Micromaps*, Daniel Carr and Chunling Zhang, George Mason University

**Tree Based Machine Learning.** **Room 1014**  
Organizer: Adele Cutler, Utah State University

1. *Random Forests Variable Selection*, Hemant Ishwaran, Cleveland Clinic
2. *Finding Structural Variants in Individual Human Genomes with Random Forests*, Jacob Michaelson, Psychiatric Genomics - Sebat Group and University of California, San Diego
3. *Probability Machines*, James D. Malley, Center for Information Technology, National Institutes of Health

3:15 pm – 3:45 pm                      Afternoon Break                      Prefunction Hall

3:45 pm – 5:30 pm                      Technical Sessions

**A Few Challenges on the Road toward Robot-Assisted ISR.** **Room 1012**  
Organizer: Barry Bodt, Army Research Laboratory

1. *Cross-country Autonomous Navigation: Trade-offs and Challenges*, Juan Pablo Gonzalez, General Dynamics Robotic Systems
2. *Using Hybrid Maps to Understand Space and Promote Common Ground*, MaryAnne Fields and Jason Owens, U.S. Army Research Laboratory and Kostas Daniilidis, University of Pennsylvania
3. *Similarity-Based Information Fusion*, Eric Heilman, Timothy P. Hanratty, U.S. Army Research Laboratory

**Reproducible Research.** **Room 1014**  
Organizer: Jürgen Symanzik, Utah State University

1. *Approaches and Barriers to Reproducible Practices in Biostatistics*, Matt Shotwell, Vanderbilt University
2. *Practical Implementation of Reproducible Research in the Madagascar Software Project*, Sergey Fomel, University of Texas
3. *Archiving Computational Research in Virtual Machines*, Sorin Mitran, University of North Carolina

5:30 pm – 7:30 pm Happy Hour at Embassy Suites

6:00 pm – 8:00 pm Board of Directors Meeting and Dinner

7:30 pm – 9:30 pm Conference Mixer at Embassy Suites

Thursday, June 2, 2011

7:30 am – 5:00 pm Registration *Bldg. C Lobby*

8:00 am – 9:45 am Technical Sessions

*Three Approaches to Large p Inference.* *Room 1012*

Organizer: David Banks, Duke University

1. *Spatial Varying Coefficient Models for Neuroimaging Data*, && Hongtu Zhu, University of North Carolina, Chapel Hill
2. *An Ordinary Differential Equation-Based Solution Path* &% Algorithm, Yichao Wu, North Carolina State University
3. *Cherry-Picking in Cluster Analysis and Regression*, David + Banks, Duke University

*Business Knowledge and Networks in SAS Data Mining.* *Room 1014*

Organizer: David Duling, SAS Institute

1. *Combining Knowledge-based and Statistical Approaches to* + Extract Sentiment from Text, Russell Albright, SAS Institute
2. *GLM models for Insurance Rate Making*, Billie Anderson, + SAS Institute
3. *Network Visualizations in Common Data Mining Analysis*, - David Duling, SAS Institute

9:45 am – 10:15 am Morning Break *Prefunction Hall*

10:15 am – 12:00 noon Technical Sessions

*Visual Inference and Uncertainty.* *Room 1012*

Organizer: Roy Welsch, MIT

1. *Data Visualization by Navigating High Dimensional Space* %& with Low Dimensional Trajectories, Wayne Oldford, University of Waterloo (Canada)
2. *Visualizing the Variability of Plots*, Rajiv S. Menjoge, Google %& and Roy E. Welsch, MIT
3. *Visual Statistical Inference for Regression Parameters*, %& Mahbubul Majumder, Heike Hofmann, and Dianne Cook, Iowa State University

**Statistical Issues in Weather and Climate Research.** **Room 1014**

Organizer: Richard Smith, SAMSI and UNC, Chapel Hill

1. *Nonparametric Spatial Models for Extreme Temperature Data*, Montse Fuentes, North Carolina State University
2. *Space-time Modeling of Climatic Trends*, Peter Craigmile, Ohio State University
3. *The Reliability of Millennial Multi-proxy Temperature Reconstructions*, Blake McShane, Northwestern University

12:00 noon – 1:30 pm      Lunch Break      *Bldg. C Café*

1:30 pm – 3:15 pm      Technical Sessions

**Celebrating the 20<sup>th</sup> Anniversary of JCGS:** **Room 1012**

**Highlights at the Interface.**

Organizer: Richard Levine, San Diego State University

1. *Cross-section and Longitudinal Penalized Functional Regression*, Jeff Goldsmith, Johns Hopkins University
2. *Dissimilarity Plots: A Visual Exploration Tool for Partitional Clustering*, Michael Hahsler, Southern Methodist University
3. *Counting Contingency Tables via Multistage Markov Chain Monte Carlo*, George Fishman, University of North Carolina, Chapel Hill.

**Maps, Nets, Optimization and Profiling:** **Room 1014**

**Data Visualization, Modeling and Interpretation in JMP.**

Organizer: Bradley Jones, JMP/SAS Institute

1. *Using Maps as a Background for Displaying Information that is Geographically Distributed*, Xan Gregg, JMP
2. *Neural Nets – Boosted and Cross-Validated*, Christopher Gotwalt, JMP
3. *Interactive Graphics for Exploring Multi-Dimensional and Multivariate Functions*, Bradley Jones, JMP

3:15 pm – 3:45 pm      Afternoon Break      *Prefunction Hall*

3:45 pm – 5:30 pm      Technical Sessions

**Algorithms for Data and Information Visualization.** **Room 1012**

Organizer: Adalbert Wilhelm, Jacobs University, Germany

1. *3-d Stereoscopic Plots: From History to R*, Juergen Symanzik, Utah State University
2. *Interactive Visualization of Multiple Time Series*, Anushka Anand, National Center for Data Mining, University of Illinois at Chicago
3. *Singular Spectrum Analysis Algorithm for Decomposing and Visualizing Time Series Data*, Michael Leonard, SAS Institute Inc.

**Statistics of Function and Shape.**

**Room 1014**

**Organizer: Michael Schimek, Medical University of Graz, Austria**

1. **Functional Varying Coefficient Models**, Damla Senturk, UCLA
2. **Smoothing Dynamic Positron Emission Tomography Time Courses Using Functional Principal Components**, Ci-Ren Jiang – UC Davis
3. **Detection of Functional Abnormalities in Brain Using Shape Analysis of Subcortical Structures**, Sebastian Kurtek, Eric Klassen, Florida State University, Zhaohua Ding, Malcolm J. Avison, Vanderbilt University, Anuj Srivastava, Florida State University

**Computing and Statistics.**

**Auditorium**

**Organizer: Edward Wegman, George Mason University**

1. **Finite User Pool Effect in Two Sample t-test of Controlled Experiments on the Web** (refereed), Shaojie Deng, Microsoft
2. **Visualizing Association Rules in Hierarchical Groups** (refereed), Michael Hahsler and Sudheer Chelluboina, Souther Methodist University
3. **The Architecture of Rc2**, E. James Harner and Mark Lilback, West Virginia University

**5:30 pm – 7:30 pm**                      **Happy Hour at Embassy Suites**

**7:00 pm – 9:00 pm**                      **Conference Banquet at Embassy Suites**

**Friday, June 3, 2010**

**7:30 am – 9:00 am**                      **Registration**                                      **Bldg. C Lobby**

**8:00 am – 9:45 am**                      **Technical Sessions**

**Business Analytics and Algorithms in Data Mining.**

**Room 1012**

**Organizer: Simon Sheather. Texas A & M University**

1. **Computing for Robust Process Engineering**, John Sall, Bradley Jones and Christopher Gotwalt, JMP/SAS Institute
2. **Two Entities and Beyond: Challenges and Examples in Social Network Analysis**, Jin-Whan Jung, Dominic Jann, Dan Kelly and John Brocklebank, SAS Institute
3. **Choosing between Logistic Regression and Classification Trees in Data Mining**, Mike Speed and Simon Sheather, Texas A&M University

**Best of Statistical Analysis and Data Mining. Room 1014**

**Organizer: Joe Verducci, Ohio State University**

1. **Symbolic Data Analysis and Statistical Analysis and Data Mining**, Lynne Billard, University of Georgia
2. **Learning Representations of Language for Greater Generalization Capacity**, Alexander Yates, Temple University
3. **Coupling Optional Polya Trees – A Bayesian Nonparametric Approach to Case-Control Studies**, Li Ma, Stanford University

**9:45 am – 10:15 am**

**Morning Break**

**Prefunction Hall**

**10:15 am – 12:00 noon**

**Technical Sessions**

**Best of Wiley Interdisciplinary Reviews:  
Computational Statistics.**

**Room 1012**

**Organizer: Yasmin Said, George Mason University**

1. **Streaming Data**, William F. Szewczyk, NSA
2. **Record Linkage**, William E. Winkler, Census Bureau
3. **Some Interdisciplinary Topics on: i) Game-Theoretical Solutions for Quantitative Cyber-Risk Estimation/ Management ii) CLOUD Computing Implementations to Compute Operational Risk using Java Programming**, Mehmet Sahinoglu, Auburn University

**Data Mining Social Data.**

**Room 1014**

**Organizer: Rida E. A. Moustafa, dMining Technology**

1. **Analysis on Racial Rates and the Effective Elements of Home Mortgage Lending Acceptance in California**, Chong Zhang, George Washington University
2. **Predicting and Explaining Potential Caravan Insurance Policy Ownership**, Haiyun Zheng, George Washington University
3. **Red Zone, Blue Zone: Discovering Parking Ticket Trends in New York City**, Samuel S. Ackerman, George Washington University
4. **Clustering TB Microarray Data for Potential Pathway Discovery and Disease Discrimination**, Laura Tipton, George Washington University